debtags.debian.net reloaded!

Feb 5, 2012
Enrico Zini (enrico@debian.org)

Introduction: debtags

Debtags is the category system for Debian packages. It started in january 2003 and is now 9 years old.

It is a faceted classification system based on an evolving, controlled vocabulary of currently 31 facets and 614 tags.

Packages are classified anonymously, and data is manually reviewed before including them in the distribution.

Debtags currently knowns of 48359 packages to tag. Of these, 28360 are managed by humans, and 19999 are still managed by robots.

(data collected febuary 1st, 2012)

Enrico Zini enrico@debian.org

Introduction: debtags.debian.net

Classification happens through a web interface designed to help with tagging as much as possible.

There are feature such as context-sensitive tagging hints, tag suggestions, and advanced tag search.

The website used to be based on CGI scripts that queried a custom C++ backend daemon which kept all the data in memory, for speed. It was hard to evolve.

I have just finished rewriting the website using Django, PostgreSQL and Xapian. The result is **debtags.debian.net**

Enrico Zini enrico@debian.org

Introduction: uses of debtags

Allow Debian to grow in size without losing structure.

Move pointless details away from package descriptions
("foo is a GTK app implemented in Python that…")

Semantic aspects of tags: hide "uninteresting" packages,
suggest packages for the current hardware, …

Domain specific package managers.

Semantically improved search functions:
http://debtags.debian.net/search
http://debtags.debian.net/search/bytag

Enrico Zini enrico@debian.org

Debtags design basics

Faceted classification: tags are grouped in omogeneous groups called 'facets', that acts as consistent points of view from which to look at the archive.

Facet examples: "role in the system", "type of user interface", "type of data it works with".

At least 7 packages per tag.

At least one use case per tag.

A tour of debtags.debian.net

- Facet list / tag list
- Knows packages from Debian and Ubuntu, soon to be extended to everything known by the Derivative Census
- Search
- Tag cloud
- Stats

Enrico Zini enrico@debian.org

A tour of debtags.debian.net

Editor
* Add/remove tags, edit patch, submit
* All tags, suggested tags, search tags
* Context-sensitive tagging hints: hardcoded, so I can provide feedback to taggers, and statistically generated, to promote tagging trends
* Anonymous submissions are tracked by cookies and aggregated together for ease of review: I don't know who you are, but you can see your tagging history!
* Experimental per-tag editor

Enrico Zini enrico@debian.org

A tour of debtags.debian.net

Patch review

- Anonymous submissions are tracked by cookes and aggregated together for ease of review: normally the edits of the same person make sense together.
- Once reviewed, they are committed to a VCS and uploaded to Debian.

Enrico Zini enrico@debian.org

A tour of debtags.debian.net

Current tag workflow
- Once manually reviewed, tags enter the archive overriding anything in debian/control.
- All versions of a package have the same tags.

Pondering possible new tag workflow:
- "Tag:" header in debian/control, binary sections.
- "Tag: foo::tag, bar::tag" in debian/control overrides all tags from debtags.debian.net
- "Tag: $DEFAULTS", or no "Tag:" header, gets tags from debtags.debian.net
- "Tag: foo::tag, bar::tag, $DEFAULTS" gets foo::tag, bar::tag, and all tags from debtags.debian.net not in facets "foo" and "bar".

Enrico Zini enrico@debian.org

A tour of debtags.debian.net

More QA: todo lists
- per maintainer
- per check
- "Getting started"

Enrico Zini enrico@debian.org

A tour of debtags.debian.net

Internals
- Exports.
- Web API.
- Maintenance scripts.
- Data sources.
- Reusable django app with the debtags website layout, site easy to theme differently for different domain names.

Enrico Zini enrico@debian.org

How to port to another distribution

Export package information in deb822 format
Export sources information in deb822 format
Popularity information is optional
Get an initial tag database via distromatch

Get in touch so we can trade submissions!

Note that you don't need all Debtags tags: even a
  selection of 5 or 6 facets will take you a long way.

Enrico Zini enrico@debian.org

Conclusion

**I believe I deserve
a rather substantial
APPLAUSE!**

Enrico Zini enrico@debian.org

Question time

Enrico Zini enrico@debian.org